EVALUATING OFF-CENTER HEAD-WORN DISPLAYS

A Dissertation
Presented to
The Academic Faculty

By

Rohan Ramakrishnan

In Partial Fulfillment
of the Requirements for the Degree
Baclelors of Science in the
School of College of Computing

Georgia Institute of Technology

December 2019

EVALUATING OFF-CENTER HEAD-WORN DISPLAYS

Approved by:

Dr. Thad Starner, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Melody Jackson
School of Interactive Computing
Georgia Institute of Technology

Date Approved: December 10, 2019

TABLE OF CONTENTS

ABSTRACT

Several studies have highlighted the advantages of using mobile augmented reality systems to assist with various tasks over traditional paper-based methods. However, these interfaces are often located in users' primary field of view which causes interference with users' vision and presents several disruptions. In this paper, a new "off-center" display type is prototyped and compared across other displays using a coloring task. Metrics such as completion time, errors, and workload are collected and used to find tradeoffs between different display types and determine their feasibility.

CHAPTER 1

INTRODUCTION AND BACKGROUND

The roots of augmented reality (AR) began in the 1960s with Sutherland's "Sword of Damocles." This early computer-based head-worn display (HWDs) was used to display 3D graphics and is one of the first HWDs created [1]. Only 30 years later would AR gain enough traction to define itself as a research field. During this time, another research field also gained popularity. This field is known as wearable computing. The goal of wearable computing is to embed sensors into everyday devices such as glasses, hoodies, and backpacks. There are three categories of augmented reality displays: Head-worn displays, handheld displays, and projection displays [1]. The focus for this study is on head-worn displays.

Head-worn displays have become more popular as their potential for assisting with tasks has been unlocked. With HWDs, users mount projectors on their heads which provides a direct integration with a users' field of vision. Objects are often connected to projectors and cameras to create dynamic interfaces that help with tasks such as order picking and navigation [2, 3]. In past projects, researchers strived to create lightweight and compact devices to attach to wearable objects that would provide a seamless transition to an augmented reality space. [4, 5]. HWDs have become more practical, lightweight devices that are rapidly gaining the interest of several manufacturers and showing promise for improving supply chain efficiency [6].

As devices become more mobile, augmented reality strives to embed graphics that provide immediate information and feedback to aid humans as they interact with the world. These innovations in mobile augmented reality have greatly increased interest and research in the

field of wearable computing.

Despite advancements to wearable technology, there are still many problems with integrating HWDs to everyday life. Their negative effects on attention and comfort are major problems that researchers are attempting to solve [6]. HWDs often interfere with users' fields of vision which can distract them while performing tasks. Visual overlap also presents serious problems. For example, if a user is operating a vehicle, they would not want navigation graphics appearing in their line of sight. Airline pilots have demonstrated that reaction time greatly decreases when on-center HWDs cause visual overlap [7].

Furthermore, users' fixation on a particular stimulus can render them blind to important changes in the environment. If a user is too fixated on the information provided by a display, they may lose information about the world. The same problem occurs if the user is fixated on their task as they will not notice important information being conveyed through the display. This issue conflicts with AR's goal of enhancing user's ability to interact with the environment.

These issues imply that displays that project interfaces off the center of vision would prove more useful as they alleviate such issues. However, studies have shown that HWDs also affect users' comfort, as looking at projected images can cause significant visual fatigue [2]. Users can gaze at a maximum of 20 degrees laterally for extended periods of time without suffering from fatigue. This tradeoff must be considered when prototyping new display types as off-center displays that project more than 20 degrees laterally cause different user experience issues.

The primary objective of this study is to discover tradeoffs when using different augmented reality displays to assist with a particular task.

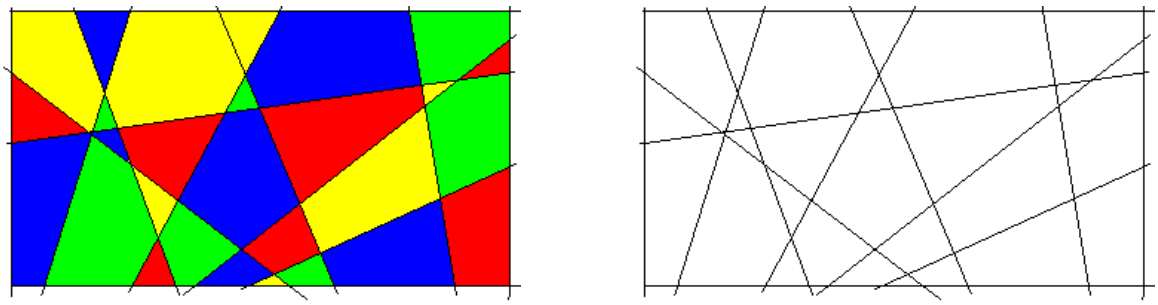
Analysis of these tradeoffs will provide further insight into the usefulness of off-center HWDs and the benefit of incorporating these types of displays in the next generation of monocular HWD design. This paper is organized as follows. Section 2 summarizes a

previous pilot study and details the hardware and methods utilized for this study. Section 3 details the results of the user study and discusses its significance. Finally, Section 4 discusses future research related to off-center displays.

CHAPTER 2

METHODOLOGY

The task was to use coloring tools provided to match a template image with a given colored image as quickly and accurately as possible.



(a) Sample colored image (b) Sample template

Figure 2.1: Sample coloring task

There has been precedent for the current research study. Three Georgia Tech graduate students used a projector and an adjustable head mount to test users' performance on a coloring task under three different display settings [8]. The results showed that off-center HWDs increased completion time and reduced error compared to the other conditions. However, the students ran the tests among their group which produced skewed results as they designed their own coloring patterns. They also only tested one example for all three conditions which causes a learning effect to affect their results [9].

This study iterated upon the pilot study to produce results across various users, collect more metrics, and collect user feedback in order to validate the pilot study's results.

(a) Static display

(b) Head-Worn display

Figure 2.2: Hardware prototypes used for display positioning

2.1 Materials

For this study, a laser beam pro C200 was used to create three display types: static on-center, static off-center, and off-center head-worn. A static on-center display is defined as a display where the projection overlays the user's field of view and cannot be moved. A static off-center display is defined as a display where the projection is located off the center of vision to the right of the user by about 10 degrees. An off-center head-worn display is defined as a display worn by the user where they are able to control the location and size of the display using head movements.

A tripod and a head mount were used to build these displays. The static displays were built

by affixing the laser display to the tripod at about 16 degrees to the vertical. The display was positioned such that the projected image was either overlaying the template or 0.25 inch to the right of the template. The head-worn display was built by affixing the laser display to the head mount as shown in Figure 2.2b. Users can adjust the device so that it fits comfortably on their heads.

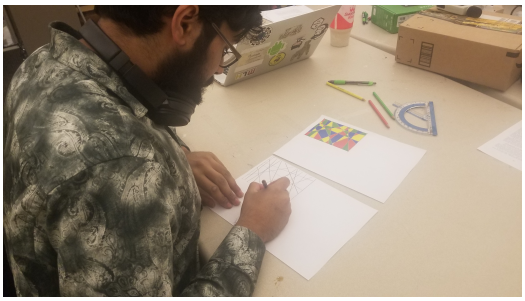
Colored images were created using GIMP. In order to reduce variance, all templates have 78 regions. The uniform number of regions ensures that the results are not skewed by using different templates. Since the laser display distorted images when projecting, it was an arduous task to print out an image that perfectly aligned with the projection. As a result the following process was used to create templates: a blank piece of paper was taped down while the static display projected the colored pattern directly onto it. Using a straightedge and pencil, the pattern was copied onto the blank paper. This process was applied to every colored image. The rest of the templates are photocopies of the hand drawn patterns.

2.2 Experimental Design

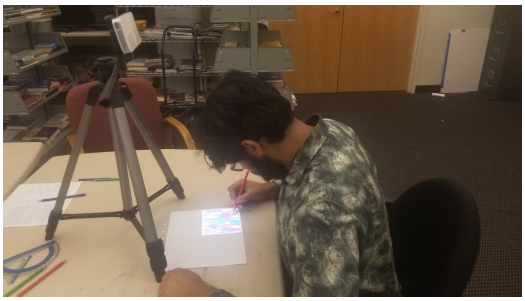
There are four conditions in this experiment. The three displays comprise three of the conditions. The fourth condition is paper-based. Users are provided the colored image and a template with the same instructions, match the colored image as quickly and accurately as possible. The paper task serves as a control to compare whether display assistance results does improve performance over using paper-based methods.

During the paper task, users could position the colored image in any position during the trial. In the static display tasks, users could not move the display.

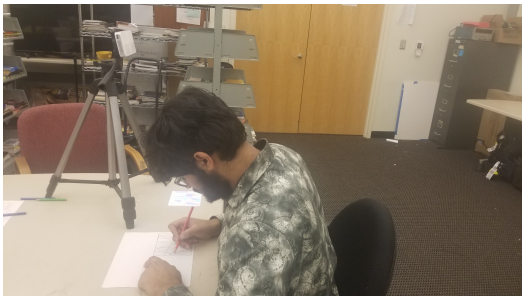
There were several aspects to consider in order to assure that the study remains valid [11]. First we must decide whether to use a within-subjects design or a between-subjects design. In a within-subjects or repeated-measures design, we test each participant under every condition (in this case, every display). In a between-subjects design, we test each



(a) Paper



(b) On-center



(c) Off-center static

(d) Off-center head-worn display

Figure 2.3: All display configurations used in the user study

participant under one condition only. There are several tradeoffs to consider based on both methodologies. Between-subjects requires more participants in order to reach a decisive conclusion, but it allows us to increase randomness and minimize bias and learning effects. Within-subjects give us more information per participant, but require us to utilize counter-balancing methods in order to prevent learning effects from having a major impact on our study.

Since the head-worn display is not as intuitive as the other conditions, a calibration task was used to help participants adjust to the head-mount. Similar to the coloring assignment, participants were provided a template while the HWD projects a colored image. This template image was missing one line compared to the original. Users identified the missing line before starting the coloring task. This initial activity reduces errors caused by the learning curve caused by using the head mount.

Originally, the images were scaled to fit the entire paper. However, users would need to rotate their heads nearly 40 degrees in order to separate the projection from the paper. To

reduce this strain, the images were scaled to take up about a quarter of the paper.

2.2.1 Counterbalancing

Learning effects occur when one participant is able to perform better in future iterations of a particular task. If one user completes a task using one type of display, even with a new coloring task, they can perform better due to learning parameters about the task they just completed. We must use counterbalancing to prevent this. However, counterbalancing using three conditions is difficult. The paper task serves to alleviate this issue and allow for a fully balanced Latin square [11].

A balanced 4x4 Latin Square is used to properly counterbalance each display configuration against each different coloring image. This method ensures that each image is tested under one of the displays positions in every possible order (1st through 4th). A balanced Latin Square also ensures that each pairing of displays to images appears before and after each other pairing at the same frequency. [9] (See Appendix A)

2.3 Metrics

The key metrics are completion time, number of errors, and workload. Errors comprise using the wrong color on a particular region and failing to color in a region. A stopwatch was used to time each participant on every condition. Participants' completed images are compared to the original and the number of errors are counted. After each trial, participants filled out a NASA Task Load Index form to gauge the perceived workload required for each display position. Additionally, participants completed a post-study survey where they list any strategies they used and add any comments on their experience.

2.3.1 NASA Task Load Index

The NASA Task Load Index (TLX) provides a measure of workload through six subjective scales. Each user fills out one TLX form to provide feedback on each display position.

For this particular assessment, users marked the spaces in between the lines, providing a minimum value of 1 and maximum value of 20 for each dimension. One notable trait is that in the performance dimension, lower scores imply better perceived performance.

2.3.2 Post-Study Survey

The post study survey requests information about strategies users used and if those strategies changed when switching between displays. Users can also comment on their experience using the different displays. Figure 2.5 shows a copy of the survey each participant completed after running through all four configurations.

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

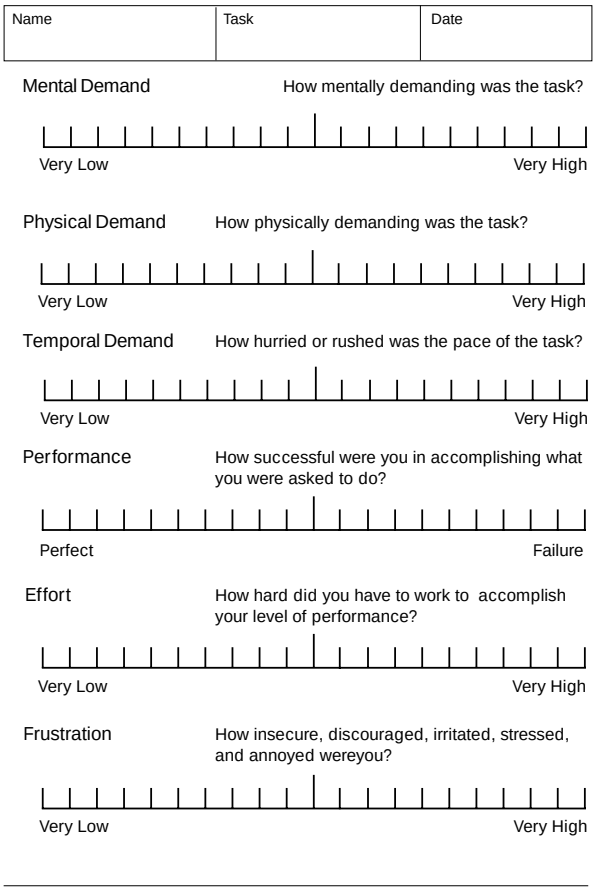


Figure 2.4: NASA TLX Form

Post-Study Questions

1. Did you use any particular strategies to complete the task?

2. Did your strategy change between different configurations?

3. How often do you participate in coloring activities?

4. Anything to share

Figure 2.5: Post-Study Survey Questions

CHAPTER 3

RESULTS

Table 3.1: Pilot Study Results.

Setup	Time (s) (Easy)	Error (Easy)	Time (s) (Hard)	Error (Hard)
Static on-center	248.5	0	531.25	1.5
Static off-center	223.25	2	462.35	3.25
HWD off-center	299.0	0.25	596	0.75

Table 3.2: Average Time and Error

Setup	Time (s)	Error
Paper	657.26	0.8125
Static on-center	589.77	1.5625
Static off-center	733.57	0.8125
HWD off-center	672.79	1.25

Table 3.3: NASA Task Load Index Values With Respect to Paper Task

Setup	Mental	Physical	Temporal	Performance	Effort	Frustration
Static on-center	-2.1875	+0.375	-0.4375	+0.5	-1.25	+0.25
Static off-center	+1.0625	+1.5	+2.125	+1.375	+1.8125	+1.5
HWD off-center	+1.5	+2.625	+0.8125	+1.125	+2.625	+2.6875

Table 3.4: Statistical Significance

Setup	p value
Static on-center vs Static off-center time	0.00395
Static on-center vs HWD time	0.0002
Static off-center vs HWD time	0.0853
Static on-center vs Static off-center error	0.07555
Static on-center vs HWD error	0.28215
Static off-center vs HWD error	0.10155

Table 3.1 details the pilot study results. The pilot study categorized coloring tasks into easy or hard images based on number of regions. For this study, Table 3.2 showcases the average completion time along with the average number of errors for each configuration

Table 3.3 displays the NASA TLX differences from the paper task. Paper is used as a reference as it is the most intuitive method to complete the coloring assignment. By calculating the differences from paper, the TLX results serve to reduce the subjective variance from different users' scores.

Table 3.4 details the one tailed p values for every condition paired against each other. This study tests for three hypotheses. Static on-center display was hypothesized to have the lowest completion time. HWD error was projected to be lower than static-on center. For the the other hypotheses, the two tailed p value should be considered since they are not a priori hypotheses. The two tailed p value is obtained by multiplying the one tailed p value by two.

CHAPTER 4

DISCUSSION

In the pilot study, errors decrease as the display moves off-center, while the time taken increases. However, the results from this recent user study shed more light on the influence of AR on this endeavor. There is a notable difference in average completion time, which rises due to the elimination of learning effects from the study. The average error also increases for most of the displays.

Off-center HWDs do tend to take slightly more time to complete due to the learning curve required to use them. But they do not seem to reduce the margin of error once learning effects are removed. The participants in the pilot study had much more experience in using the head-worn displays so they were able to use their experience to their advantage. While the calibration task is useful to allow new users to adjust to the new experience, it was not sufficient and new participants are at a disadvantage when compared to the pilot study's participants.

While the weight of the head-worn device can only be fixed by prototyping, the learning curve issue can be addressed with more a rigorous study design. If studies spanned multiple sessions, where the initial sessions focused on encouraging the user to experiment with each display configuration on simple tasks such as the missing line test. Naturally this method requires tuning to prevent learning effects from skewing the results, but this modification can help offset the learning curve presented by the off-center HWD.

Furthermore, off-center static error decreases significantly. This result highlights the potential for off-center AR to reduce errors when used to assist with basic tasks. The time taken increases which means that participants spent more time correcting errors and ensur-

ing their accuracy. While the paper and off-center static conditions tie for the lowest error, AR interfaces that provide immediate feedback is a possible modification. With paper, such real time feedback is impossible.

The on-center task incurred the most errors since users were forced to switch between stimuli. It was also difficult to determine which regions had been colored in beforehand and which ones remained. This disadvantage greatly increased the number of incorrectly colored regions. This particular trend is consistent between both studies.

Using $\alpha = 0.05$ as a baseline, there are only two statistically significant hypotheses defined by $p < \alpha$. This means that the hypothesis that on-center completion times are, on average, lower than those of off-center static displays and off-center HWDs is true with reasonable confidence. The other hypotheses with higher p values are inconclusive and require more testing to validate.

More rigorous studies across more complex AR assisted activities are required to confirm these results, but off-center HWDs still show great potential in advancing the field of both AR and wearable computing.

4.1 NASA Task Load Index Responses

While the NASA TLX is inherently subjective, the results do reveal users' perceived effort and performance between each condition. The off-center HWD took the most mental and physical demand and caused the most frustration out of each task. This trend is due to the learning curve required. Despite the calibration task, the discomfort of the prototype outweighed the advantage of control over the projection. With a lighter HWD, the advantages would outshine the detriment of learning the projection.

4.2 Survey Responses

According to the surveys, the most commons strategy was to pick one color and fill in all the regions corresponding to that color before moving onto the next till all regions were filled in. Some users specified starting with dark colors and filling in lighter colors at the end.

Some users also marked regions with a quick scribble based on the image, then focused on coloring in the regions and ignoring the projected image once all regions were marked.

Six of the sixteen users reported changing strategies between displays which either implies that some strategies were suboptimal for different configurations.

Many users commented on the discomfort of using the head-worn display due to the combined weight of the head-mount and the laser display as well which likely factored into their error and completion time.

4.3 Sources of Error

Before analyzing the sources of error, there are aspects of the task itself that affect the performance between displays. The coloring task is a static task which favors the static conditions. HWDs are particularly suited for mobile tasks where users must move around while completing a given assignment. But since such tasks would so heavily favor HWDs, a task that did not require significant motion was selected instead to highlight trends between the different displays. Despite this disadvantage, the pilot study results implies that experience plays a large factor in using the HWD as opposed to the other displays that are more intuitive.

HWDs are often built to be lightweight so as to reduce the physical strain of wearing it. This HWD prototype unfortunately was particularly heavy which affected both completion time and error. While the calibration test attempted to offset this, it is a shorter test than

the coloring test. The weight and discomfort from using the display worsened performance over longer periods of time due to the sustained effort in utilizing the tool.

It is important to note that when transcribing the images from a projection to paper, some very small gaps appeared which did not correspond to any region in the image. These gaps mapped to grey or black pixels. Some users did fill in these gaps while others ignored them. These distorted gaps were not factored into error calculation and were so small that the completion time is not significantly affected.

Additionally, the colored images assigned to participants did not undergo a rigorous graph coloring procedure. As a result many adjacent regions shared the same color. This issue can lead to perceptual issues and result in more errors. A graph coloring procedure is outlined below.

4.3.1 Graph Coloring

Vertex coloring is a common algorithms problem where colors are assigned to elements of a graph such that no two colors are adjacent. Vertex coloring algorithms can be useful in designing a more precise coloring task. By defining each region as a vertex and each edge based on the adjacency of regions, a graph can be created and run through a vertex coloring algorithm such that each region is assigned a distinct color from its neighbors. This procedure may requires additional colors depending on the graph structure.

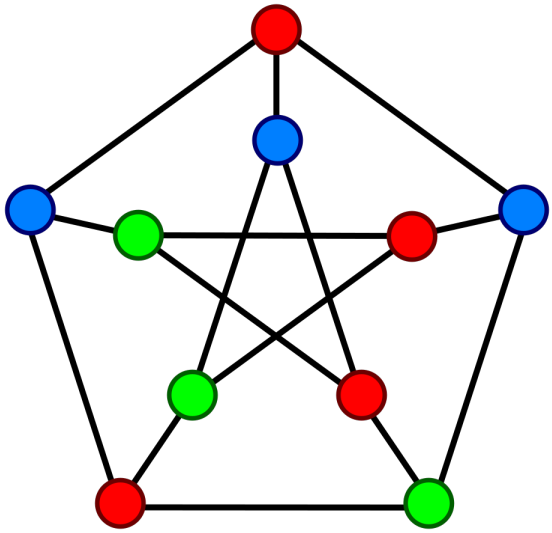


Figure 4.1: Graph coloring example

CHAPTER 5

CONCLUSION AND FUTURE WORK

Designing a completely new HWD that makes use of an off-center display is a significant investment, and this study cannot confidently conclude that this investment would be worth the undertaking. Further studies must be conducted to gauge the value of off-center HWDs and their potential utility.

There are several variations of this particular study that can be run to discover new factors related to display positioning. Adding more colors to this task would force users to observe the display more often. Similarly, having more sparse colored regions would reduce time but also encourage users to observe the displayed image more often so they do not compromise their accuracy. Both these variations attempt to identify how tasks that require more attention on the interface would be affected by different display positions.

Other metrics may also be useful for this particular study. Color quality can be defined by how well users restrict colors to their respective regions. Significant overlap between regions implies poorer color quality. This metric can be measured by presenting all four colored papers to a third party panel of judges. The judges rank the papers from best to worst. This provide insight into how well participants can complete this assignment without overlapping colors between regions under all conditions.

Though this particular study could be reapplied with different parameters, other tasks should be run through different display types in order to identify which display positions are best suited for different tasks. New tasks would also provide insight into the properties of activities that make certain HWDs better for AR assisted task completion. Another interesting task uses a children's sticker book and a display to create a different activity. The

display would project an image from the book that is uncolored and labeled with a number. This number corresponds to a labeled sticker in the book. Users would have to map that display to the sticker and manually apply this sticker to their uncolored image and match the correct sticker to each region as quickly as possible. The current methodology can still be applied to this new task, but users would likely spend more time and effort identifying and using stickers. Thus, this experiment would measure task performance where users spend more time visually attending to real world stimuli as opposed to the display's interface. Additionally, error would likely be 0 due to the distinct stickers. User's would be able to identify mistakes based on the shape of the projection. Therefore, this task emphasizes completion time.

There are many tasks where static on-center interfaces are infeasible. Such tasks can still be run through static off-center interfaces and off-center HWDs to compare whether portability and control over the display are more or less useful for different activities.

Varying the types of interfaces used would also provide insight into the utility of off-center HWDs. The interface used in the aforementioned study is a static image with no feedback mechanism. New, interactive interfaces can also be tested using these methodologies. Furthermore, interfaces that provide direct feedback on user error should be tested across these different displays. By considering all these factors, future studies can make a cogent case for or against the adoption and production of off-center HWDs.

Appendices

APPENDIX A

INDIVIDUAL RESULTS

Table A.1: Individual Time and Error

Participant	Paper Time (s)	On-center Time (s)	Off-center Static Time (s)	HWD Time (s)
1	615.94	554.59	526.34	560.68
2	455.2	335.35	415.94	378.18
3	428.53	305.5	635.71	563.37
4	728.86	580.87	972.83	734.5
5	1155.04	1002.88	896.22	947.45
6	657.95	621.05	1089.37	672
7	1039.79	1110.25	1388.12	1181
8	388.72	326.03	415.74	369.03
9	495.73	257	444.04	368.58
10	437.91	689.48	509.97	743.34
11	974.18	972.95	949.48	1136.93
12	510.69	543.94	563.9	591.4
13	535.43	505.67	589.61	655.02
14	500.04	448.76	579.73	491.23
15	817.12	746.88	1119.83	833.13
16	775.05	435.05	640.36	538.73

Table A.2: Individual Time and Error

Participant	Paper Error	On-center Error	Off-center Static Error	HWD Error
1	0	1	2	2
2	1	3	1	2
3	0	1	2	3
4	0	1	1	0
5	0	0	0	0
6	2	4	0	0
7	0	0	0	0
8	1	1	3	1
9	5	1	4	5
10	1	2	0	2
11	1	4	0	0
12	0	0	0	1
13	1	1	0	0
14	0	2	0	0
15	0	2	0	0
16	1	2	0	4

Table A.3: NASA Task Load Index Values for Paper Task

Participant	Mental	Physical	Temporal	Performance	Effort	Frustration
1	6	7	5	1	8	6
2	14	7	11	2	10	11
3	3	1	5	1	2	2
4	7	6	7	4	6	4
5	4	11	14	7	5	3
6	1	2	1	2	2	2
7	5	6	8	6	5	4
8	8	8	13	1	6	1
9	18	18	7	3	16	8
10	2	2	10	2	2	2
11	9	5	10	3	10	8
12	4	6	4	2	5	3
13	6	4	10	4	8	2
14	11	11	7	4	5	5
15	7	7	9	1	8	8
16	1	1	13	2	1	1

Table A.4: NASA Task Load Index Values for On-Center Task

Participant	Mental	Physical	Temporal	Performance	Effort	Frustration
1	2	6	2	1	3	2
2	11	7	14	4	9	13
3	3	2	2	3	1	2
4	5	5	6	5	4	5
5	3	10	13	4	4	3
6	4	9	10	2	7	7
7	3	7	10	5	3	2
8	8	8	13	1	4	1
9	2	11	8	3	4	3
10	5	8	2	5	7	7
11	3	7	6	1	11	6
12	2	7	3	2	5	2
13	3	5	10	7	2	4
14	10	11	10	5	5	4
15	6	4	2	3	5	12
16	1	1	16	2	5	1

Table A.5: NASA Task Load Index Values for Off-Center Static Task

Participant	Mental	Physical	Temporal	Performance	Effort	Frustration
1	3	5	3	1	4	2
2	14	8	14	2	16	9
3	6	2	10	2	5	2
4	8	6	5	4	5	7
5	2	11	11	3	4	3
6	6	11	8	3	7	8
7	5	10	12	5	12	5
8	8	8	13	2	12	3
9	20	20	20	13	19	19
10	3	9	7	3	3	5
11	15	11	14	18	16	6
12	6	5	5	2	5	2
13	8	4	11	3	7	7
14	12	11	10	4	7	8
15	6	4	10	1	5	7
16	1	1	15	1	1	1

Table A.6: NASA Task Load Index Values for Off-Center HWD Task

Participant	Mental	Physical	Temporal	Performance	Effort	Frustration
1	6	5	4	1	4	2
2	11	9	13	3	8	10
3	7	4	7	4	6	5
4	5	5	7	4	6	2
5	3	14	13	4	6	8
6	3	5	5	3	4	5
7	15	14	5	5	15	15
8	9	5	12	1	13	1
9	19	18	16	12	15	17
10	13	13	13	7	16	11
11	9	20	15	5	14	4
12	6	11	3	3	10	13
13	8	0 (unreported)	8	5	10	10
14	11	12	6	3	6	7
15	2	8	4	1	4	2
16	3	1	16	2	4	1

Table A.7: Latin Square Ordering

Participant	first	second	third	fourth
1	A1	B2	D3	C4
2	B3	C1	A2	D4
3	C3	D2	B1	A4
4	D1	A3	C2	B4
5	A2	B3	D4	C1
6	B4	C2	A3	D1
7	C4	D3	B2	A1
8	D2	A4	C3	B1
9	A3	B4	D1	C2
10	B1	C3	A4	D2
11	C1	D4	B3	A2
12	D3	A1	C4	B2
13	A4	B1	D2	C3
14	B2	C4	A1	D3
15	C2	D1	B4	A3
16	D4	A2	C1	B3

APPENDIX B

DETAILED EXPERIMENT INSTRUCTIONS

B.1 Resources

The hardware to recreate this experiment is located at TSRB lab 243. Contact Blue Lin (glin39@gatech.edu) or Thad Starner for more information and for access to the Google drive with all resources.

At least one C200 Laser Display (https://www.amazon.com/Laser-Beam-Pro-Resolution-Rechargeable/dp/B01NAPMAQ6) is required to run this experiment. This display has a hole on the back where a 1/4"-20 screw fits in. Using the screw and a drill, this display can be affixed to any other hardware.

The current coloring tasks are located in a google drive. I highly recommend creating new coloring templates with a specified number of regions using a graph coloring algorithm. Another recommendation is to make the tasks color blind friendly.

You will also need coloring utensils (this study used colored pencils). I recommend having at least 2 spares per color in case of breaking. If you use colored pencils be sure to provide a pencil sharpener. Tell participants to grab a new coloring utensil if one breaks and handle sharpening between tasks.

B.2 Running the Experiment

Start by providing participants with a copy of the consent form and show them the script video (assuming you're only reproducing this study). Afterwards, give them the NASA TLX instruction sheet (I personally used pages 12 and 13) and allow them to read over

each measure. Make sure to stress to participants that they are aiming to be as quick and accurate as possible when matching the image.

Use the 4x4 balanced Latin Square ordering for each participant [9]. A diagram is located in Appendix A. Map each letter to a display type (paper, on-center, off-center static, off-center HWD) and each number to a coloring image.

Tape the template down for all displays. For paper, participants may adjust the paper as they would like. For static on-center, line up the projection as best as possible with the coloring template before taping down the template. Note that if you need more precise templates, it's best to tape a blank piece of paper and use a straightedge to draw each line in the image. Once complete, photocopy the hand-drawn template and keep the original for future copies.

For off-center static, I moved the display so that the projection was about 0.25 inch from the paper. Users are instructed to not translate and rely on head rotations to observe the projected image. Note that when using the images provided that they map to the top left forcing users to rotate their heads more. Using photo editing software such as GIMP, the modified image can be modified to snap the image to the top right instead (I recommend this as it may yield more conclusive results).

For off-center head-worn display, the calibration task outlined in Chapter 2 is useful to provide with some experience before starting the coloring activity. It may be useful to create multiple more missing line tests using different templates.

Start a stopwatch once users start coloring and stop it once they are completed. Provide a NASA TLX form immediately after completing every activity before moving onto the next. Make sure users do not have access to their previous TLX forms for reference (this is how this study handled that. Future studies might change this aspect). At the end, provide the user survey and collect all documents. Count the number of errors for each task and note them down in a spreadsheet.

When reporting results for the NASA TLX, make sure to always report the differences between a baseline task (in this case, the paper task serves as the TLX baseline). See Table 3.3 for an example.

REFERENCES

[1] R. Azuma *et al.*, "Recent Advances in Augmented Reality," 2001.

[2] C. Thomas *et al.*, "RF-Pick: Comparing Order Picking Using a HUD with Wearable RFID Verification to Traditional Pick Methods," 2017.

[3] S. Feiner *et al.*, "Invisible AR HWD: Using Non Line-of-Sight Displays to Perform Real World Tasks," 2017.

[4] T. Hollerer *et al.*, "Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system," 1999.

[5] P. Mistry, P. Maes, and L. Chang, "WUW - Wear Ur World - A Wearable Gestural Interface," 1935.

[6] M. Haynes, "LATERAL POSITIONING OF A MONOCULAR HEAD-WORN DISPLAY," PhD thesis, 2017.

[7] T. Starner, "A User-Centered Approach to On- the-go HWD Design.," 2018.

[8] M. Aki, A. Khan, and H. Wang, "Invisible AR HWD: Using Non Line-of-Sight Displays to Perform Real World Tasks," 2017.

[9] *Within-subjects vs. Between-subjects Designs: Which to Use?* `https://www.yorku.ca/mack/RN-Counterbalancing.html`.